Microarray Techniques to Analyze Copy-Number Alterations in Genomic DNA: Array Comparative Genomic Hybridization and Single-Nucleotide Polymorphism Array

Károly Szuhai[1] and Maarten Vermeer[2]

Journal of Investigative Dermatology (2015) **135**, e37. doi:10.1038/jid.2015.308

INTRODUCTION

Tumor cells often exhibit alterations in chromosome number and/or structure. Studies of cells in metaphase enable detection of these alterations. However, archived samples cannot be used for this approach, and access to proliferating tumor cells for metaphase analysis can be difficult, especially when attempting to culture cells from solid tumors. Comparative genomic hybridization (CGH), described in 1992, allows comparison between tumor and normal cell genomes by differentially labeling each and cohybridizing them on slides containing normal metaphase cells. This technique makes it possible to detect common regions of genomic deletion and amplification within archived tumor samples. However, initially the resolution was limited to large genomic changes visible on metaphase chromosomes. To improve the resolution, human metaphase chromosomes were replaced by an array or matrix on a chip of bacterial artificial chromosome (BAC) clones generated using human genome fragments. These were later replaced by commercially available 25–70 base pair–long oligonucleotides, which provide high accuracy and reproducibility. Here, we explain these approaches, known as array CGH or single-nucleotide polymorphism array (SNP array), highlighting their benefits, limitations, and potential applications and discussing their future in view of now widely available next-generation sequencing (NGS) techniques.

Array CGH and SNP arrays have been used extensively to analyze copy-number alterations in tumor DNA. Both techniques provide a genome-wide screening tool to identify deletions and amplifications, and SNP arrays also allow allelotyping. Array approaches are used to search for causative constitutional (germline) and acquired genomic alterations in carcinogenesis. These techniques were utilized to identify recurrent copy-number alterations in tumors and led to the identification of novel oncogenes and tumor suppressor genes. Results from array-CGH studies have been important for the subclassification of dermatological malignancies and identification of novel potential prognostic markers and genomic alterations involved in dermatological disease pathogenesis.

WHAT GENOMIC ARRAY DOES

- Is a genome-wide screening tool.
- Interrogates hundreds of thousands to millions of probes in a single experiment.
- Detects tumor-associated genomic changes with copy-number alterations (gains/amplifications and losses/deletions).
- Probes designed for single-nucleotide polymorphism loci allow allelotyping.

LIMITATIONS

- Requires large, up-front investment in equipment (hybridization, imaging of slides).
- Balanced rearrangements are not detected.
- Single-nucleotide changes (pathogenic mutations) will not be detected.
- Studies a mixed pool of cells, so there are limitations detecting changes in cases of normal cell infiltration and high tumor heterogeneity.
- Does not provide data on expression levels of genes involved in the gains or amplifications.

HOW ARRAY TECHNIQUES WORK

In array approaches, probes corresponding to genomic regions are immobilized onto a glass surface, forming the detection chip. Isolated and fluorescently labeled test DNA is then competitively hybridized to these chips and any unbound DNA is stringently removed by washing. Fluorescence intensities are scanned by an imager and correspond to the copy number of each interrogated loci (probe). In contrast to fluorescence *in situ* hybridization (FISH), which generates data about genomic loci at a single-cell level, genomic array approaches provide information about pools of cells, albeit without morphological context.

ARRAY CGH VERSUS SNP ARRAY

For array CGH, the labeled test sample is mixed with a differently labeled reference DNA sample that typically consists of a pool of DNA isolated from several healthy donors. The use of a normal reference DNA sample allows easy normalization within one experiment (Figure 1, left).

[1]*Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands and* [2]*Department of Dermatology, Leiden University Medical Center, Leiden, The Netherlands*

Correspondence: Károly Szuhai, Department of Molecular Cell Biology, Leiden University Medical Center, Eithovenweg 20, 2333 AL, Leiden, The Netherlands. E-mail: k.szuhai@lumc.nl

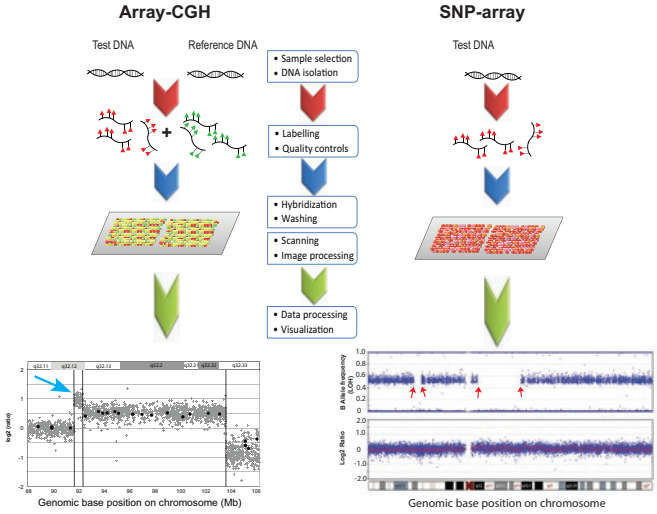

Figure 1. Overview of genomic array-based procedures. The left panel shows array-comparative genomic hybridization (array CGH), whereas the right panel shows single-nucleotide polymorphism (SNP) array. Array CGH utilizes differentially labeled test and reference DNAs, whereas SNP array uses labeled single test samples. In both cases, samples are mixed with repeat suppressor C_{0t}-1 DNA and hybridized to detection chips. After hybridization and stringent washing, array chips are scanned by a fluorescence laser scanner and intensity values are further processed. Resulting data files are analyzed for changes in chromosomes and DNA base positions. The resolution of the array is largely determined by the reporter probe density. The lower left panel shows superimposed results using either an ~3.5-kb bacterial artificial chromosome (BAC) probe array (black dots) or a 244-kb oligonucleotide probe array (gray dots). The blue arrowhead indicates a small triplicated region undetected by the BAC array. The lower right panel shows results from a SNP array showing the copy number of the displayed chromosome (lower inset), indicating no copy-number alteration. In the upper inset, B-allele frequency is displayed. Because most of the SNP should be heterozygous, here the allele frequency should be expected to be 0.5. Loss of heterozygosity (LOH) would lead to either the lack of B alleles (0) or two B alleles (1.0). The red arrowheads mark large regions of LOH without corresponding copy-number alterations, a phenomenon known as copy number–neutral LOH.

For SNP array, reporter oligonucleotides, also known as allele-specific oligonucleotides, are designed to detect known single-nucleotide differences, selected from all known variants for the interrogated sequence. In an SNP-array experiment, a single test sample is hybridized to a chip and detected with a single label (Figure 1, right). With bioinformatics algorithms, SNPs can be identified and the respective copy number can be deciphered. Detection of both loss of heterozygosity (LOH) and copy-neutral anomalies, such as uniparental disomy or homozygosity, becomes feasible using SNP arrays (Figure 1, right). Use of DNA derived from both tumor and normal tissue from the same patient is necessary to draw conclusions about tumor-associated changes.

Finally, normalized fluorescence ratios are then ordered based on their genomic position and displayed as a molecular karyogram (Figure 2a).

ADVANCES IN ARRAY APPROACHES

Multitarget detection and wide range in copy-number detection capability

In contrast to targeted approaches such as FISH or quantitative, locus-specific PCR, an array experiment allows simultaneous testing of thousands of loci, allowing genome-wide

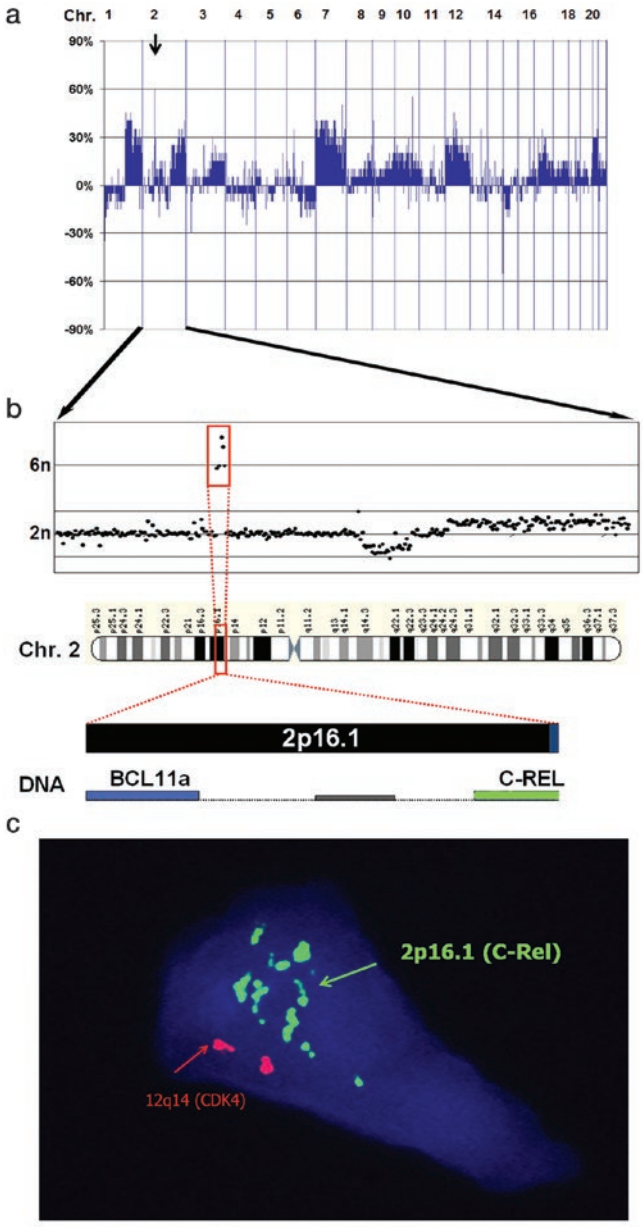

Figure 2. Array CGH identified a high level of amplification and homozygous losses in primary cutaneous large B-cell lymphomas. (**a**) Genome-wide overview of multiple primary cutaneous large B-cell lymphoma cases. (**b**) The most frequent high level of amplification involved the chromosome 2p16.1 region where *BCL11A* and *REL* genes reside. (**c**) Interphase: fluorescence *in situ* hybridization (FISH) verification of the *REL* amplification using REL (green) and CDK4 (red) region-specific probes. The numerous loci of green signal correspond to the amplification detected by array CGH. Reprinted with permission from Dijkman *et al.* (2006).

screening of copy-number alterations. Further, the relative number of altered copies of loci can be determined, ranging from a homozygous deletion to a high number of amplifications, since the fluorescence signal intensity is linearly proportional to the number of tested molecules.

Heterogeneous populations

Rigorous quality measures allow detection of genomic changes present in low frequency (8–10%) in a heteroge-

neous population using array approaches. Examples include mosaicism found when studying constitutional genetics, intratumor heterogeneity, or tumor samples infiltrated with normal cells (Szuhai *et al*., 2011).

LIMITATIONS OF ARRAY TECHNOLOGIES

Resolution

The resolution of array technologies is determined by the probe density. A higher density allows more precise detection of small genomic changes, but may increase cost per analyzed sample and background, such as detection of chromosomal copy-number variations with unknown clinical significance (Figure 1, left) (Knijnenburg *et al*., 2007). At present, the costs for NGS tests are similar to costs for array-based experiments, with several advantages of NGS over genomic arrays, including the ability to identify balanced genomic rearrangements and small mutations. For example, exome sequencing combined with the proper analysis algorithm generates a genome-wide copy-number profile, allelotyping, and details on single-nucleotide changes (Grada and Weinbrecht, 2013). Increased use of targeted sequencing-based approaches could further decrease test-related costs and detect all relevant gene mutations, copy-number changes, and translocation breakpoints (Cheng *et al*., 2015).

Spatial organization

Detecting genetic copy-number changes will not reveal information about the spatial organization of the involved chromosomes. For example, an unbalanced translocation will appear as a loss and gain of the involved chromosomal regions, but nothing about their physical association will be revealed. Similarly, balanced genomic rearrangements including balanced translocations among chromosomes, chromosomal insertions, and intrachromosomal changes like inversions cannot be detected using array approaches.

Sample availability

DNA isolated from fresh-frozen tissues is preferred for array techniques because DNA isolated from formalin-fixed, paraffin-embedded tissue blocks may have undergone degradation, leading to nonuniform sample labeling with fluorescent nucleotides.

For accurate detection of genomic alterations, samples with as high tumor content as possible (>50% or, ideally, 80–90% or higher) should be selected. The detection sensitivity depends not only on a high tumor content, but also on the ploidy of tumor cells. Because of the normalization process, samples with tetraploid DNA content will show only half the dynamic range for alterations compared with those with a diploid DNA content.

DATA INTERPRETATION

For array CGH, where the ratio of the test and reference samples is taken from the same hybridization reaction, internal normalization happens immediately, assuming that no great quality differences were present between the test and reference DNA. For SNP arrays, each measurement is compared and normalized to a previously generated external reference. Fluctuations in actual test quality may therefore have a stronger effect.

Various statistical approaches can then be applied to normalize the data across the whole genome. The obtained ratios are often displayed as Log_2 values to help detect gains and losses. Results are displayed according to chromosome and genomic positions (Figure 1).

Validation

Resulting data can be validated using other techniques, such as quantitative real-time PCR or FISH (Figure 2). Multiplex ligation-based probe assays can be used to verify up to 40 loci in a single reaction. However, for novel findings new probes may need to be designed, and FISH probes may not have the resolution to confirm results from array-based approaches. For example, gains and deletions of 5–10 kb can be readily detected by array approaches, whereas most FISH probes generated from BACs are between 80 and 200 kb. Gains or duplications, especially small tandem duplications, may also not be detectable using FISH probes.

USE OF GENOMIC ARRAYS IN DERMATOLOGY

Genomic array approaches have made major contributions to understanding dermatological cancers. In melanoma, array-CGH approaches revealed that increased chromosomal instability is associated with poor clinical outcome (Hirsch *et al*., 2013). In addition, array-CGH studies uncovered homozygous deletions of the tumor suppressor genes *CDKN2A* and *PTEN* and amplifications of oncogenes *MITF, CCND1*, and *MDM2*, suggesting that these copy-number alterations play a role in the pathogenesis of melanoma since these genes regulate increased cell cycle progression and resistance to apoptosis (Stark and Hayward, 2007).

Extensive genetic instability and recurrent copy-number alterations were found in Sézary syndrome (SS) using array CGH, including gain of *MYC*, loss of the MYC antagonists *MXI1* and *MNT*, loss of *TP53*, and gain of *STAT3/STAT5* and IL-2-receptor genes. These results suggest that increased MYC expression combined with impaired MYC-induced apoptosis and increased STAT signaling play a role in the pathogenesis of SS (Vermeer *et al*., 2008).

SS is often considered a leukemic phase of mycosis fungoides (MF). However, substantial genomic differences between SS and MF were found using array-CGH and gene expression microarrays, suggesting a distinct molecular pathogenesis. Furthermore, a high number of DNA alterations, specifically gains of 8q24.21 and losses of 9p21.3 and 10q26qter, were found to be associated with poor prognosis in MF (Salgado *et al*., 2010; van Doorn *et al*., 2009).

Array-CGH and gene expression studies were important in supporting the subclassification of primary cutaneous large B-cell lymphomas into primary cutaneous follicle center lymphoma (PCFCL) and primary cutaneous large B-cell lymphoma (PCLBCL), leg type. This distinction is clinically important because PCFCL has a good progno-

sis and can be treated with radiotherapy, whereas PCLBCL, leg type, has a poor prognosis (5-year survival of 40%) and should be treated with chemotherapy. A high level of amplification of the chromosome 2p16.1 region was observed in 60% of PCFCL cases, involving *BCL11A* and *REL* genes (Figure 2). Homozygous *CDKN2A* deletion is associated with poor prognosis in PCLBCL, leg type (Dijkman *et al.*, 2006; Senff *et al.*, 2009). These studies highlight how array-CGH data have been used in the subclassification of dermatological malignancies and have identified potential prognostic markers and genomic alterations that may play a role in the pathogenesis of these diseases.

SUMMARY AND FUTURE PROSPECTS

Genomic array testing allows detection of genetic copy-number changes and, in the case of SNP array, allelic changes. Results from genomic array tests have been important in the identification of underlying copy-number alterations in microdeletion/microduplication syndromes and in characterizing acquired genetic alterations in malignancies. Collected, publicly accessible array profiles could serve as references to gain a broader understanding of the genetic etiologies underlying disease. However, in our view, genomic array techniques will likely soon be overtaken by NGS-based approaches to uncover unknown genomic changes or generate novel diagnostic tests.

CONFLICT OF INTEREST

The authors state no conflict of interest.

CME ACCREDITATION

This activity has been planned and implemented in accordance with the Essential Areas and Policies of the Accreditation Council for Continuing Medical Education through the joint sponsorship of the Duke University School of Medicine and Society for Investigative Dermatology. The Duke University School of Medicine is accredited by the ACCME to provide continuing medical education for physicians. To participate in the CME activity, follow the link provided. Physicians should only claim credit commensurate with the extent of their participation in the activity.

To take the online quiz, follow the link below:
http://continuingeducation.dcri.duke.edu/research-techniques-made-simple-journal-based-cme-rtms

SUPPLEMENTARY MATERIAL

A PowerPoint slide presentation appropriate for journal club or other teaching exercises is available at http://dx.doi.org/10.1038/jid.2015.308.

REFERENCES

Cheng DT, Mitchell TN, Zehir A *et al.* (2015) Memorial Sloan Kettering–Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn* 17:251–64

Dijkman R, Tensen CP, Jordanova ES *et al.* (2006) Array-based comparative genomic hybridization analysis reveals recurrent chromosomal alterations and prognostic parameters in primary cutaneous large B-cell lymphoma. *J Clin Oncol* 24:296–305

Grada A, Weinbrecht K (2013) Next-generation sequencing: methodology and application. *J Invest Dermatol* 133:e11

Hirsch D, Kemmerling R, Davis S *et al.* (2013) Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma. *Cancer Res* 73:1454–60

Knijnenburg J, van Haeringen A, Hansson KB *et al.* (2007) Ring chromosome formation as a novel escape mechanism in patients with inverted duplication and terminal deletion. *Eur J Hum Genet* 15:548–55

Salgado R, Servitje O, Gallardo F *et al.* (2010) Oligonucleotide array-CGH identifies genomic subgroups and prognostic markers for tumor stage mycosis fungoides. J Invest Dermatol 130:1126–35

Senff NJ, Zoutman WH, Vermeer MH *et al.* (2009) Fine-mapping chromosomal loss at 9p21: correlation with prognosis in primary cutaneous diffuse large B-cell lymphoma, leg type. *J Invest Dermatol* 129:1149–55

QUESTIONS

This article has been approved for 1 hour of Category 1 CME credit. To take the quiz, with or without CME credit, follow the link under the "CME ACCREDITATION" heading.

For each question, more than one answer may be correct.

1. Comparative genomic hybridization is a tool that is suited to

A. Detect copy-number alterations.
B. Detect translocations.
C. Give information about gene expression.

2. A genomic array-CGH profile without any detectable alteration indicates

A. The lack of tumor cells.
B. Copy number–neutral LOH.
C. Balanced rearrangements.
D. Probe density that is too low.

3. Array CGH is preferably performed using DNA obtained from

A. Frozen tissue with low tumor content.
B. Paraffin-embedded tissue with low tumor content.
C. Frozen tissue with high tumor content.
D. Paraffin-embedded tissue with high tumor content.

4. Major advances in array approaches do *not* include the following:

A. Multiple genetic loci are investigated in one experiment.
B. A wide dynamic range in copy-number estimation allows for detection of gains and losses.
C. Array CGH can be performed on heterogeneous cell populations.
D. Information on spatial organization of the involved chromosomes.

5. The resolution for the detection of copy-number alterations depends largely on

A. The probe density on a chip.
B. The software that is used.
C. The DNA quality of the samples.
D. The amount of available sample DNA.

Stark M, Hayward N (2007) Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Res* 67:2632–42

Szuhai K, Jennes I, de Jong D *et al*. (2011) Tiling resolution array-CGH shows that somatic mosaic deletion of the EXT gene is causative in EXT gene mutation negative multiple osteochondromas patients. *Hum Mutat* 32:E2036–49

van Doorn R, van Kester MS, Dijkman R *et al*. (2009) Oncogenomic analysis of mycosis fungoides reveals major differences with Sezary syndrome. *Blood* 113:127–36

Vermeer MH, van Doorn R, Dijkman R *et al*. (2008) Novel and highly recurrent chromosomal alterations in Sezary syndrome. *Cancer Res* 68:2689–98